Econometric Analysis

Ryuichi Tanaka

Ch. 8 System Equation by IV

Example 8.1 only equilibrium is observed (labor supply and demand functions)

Labor supply function $h^s(w) = \gamma_1 w + \mathbf{z}_1\boldsymbol{\delta}_1 + u_1$ z : controls (educ, age, married)

Wage function $w^o(h) = \gamma_2 h + \mathbf{z}_2\boldsymbol{\delta}_2 + u_2$ wage for workers with h hours work

We can observe only equilibrium (h,w) satisfying the following two equations

$$h = \gamma_1 w + \mathbf{z}_1\boldsymbol{\delta}_1 + u_1 \quad \text{and} \quad w = \gamma_2 h + \mathbf{z}_2\boldsymbol{\delta}_2 + u_2$$

In $h = \gamma_1 w + \mathbf{z}_1\boldsymbol{\delta}_1 + u_1$, u_1 **and** w **are correlated**

In $w = \gamma_2 h + \mathbf{z}_2\boldsymbol{\delta}_2 + u_2$, u_2 **and** h **are correlated => IV estimation is needed!**

IV estimation of system equation is more efficient than IV equation by equation

8.1 Introduction and Examples

Data : $\{(\mathbf{X_i}, \mathbf{y_i}) : i = 1,2,\cdots,N\}$, $\mathbf{X_i} : G \times K$, $\mathbf{y_i} : G \times 1$

Model: $\mathbf{y}_i = \mathbf{X_i}\boldsymbol{\beta} + \mathbf{u}_i$ (system equations)

ASOLS.1. $E(\mathbf{X_i}'\mathbf{u}_i) = \mathbf{0}$ => POLS is consistent

If not satisfied, need to estimate by IV

It is common approach to estimate by GMM (by Wooldridge)

Generalized Method of Moment

8.2 A General Linear System of Equation

Model : $\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{u}_i$, y: Gx1, X: GxK (In panel, **G=T**)

$$\mathbf{y}_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iG} \end{pmatrix},\ \mathbf{X}_i = \begin{pmatrix} \mathbf{x}_{i1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_{i2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{x}_{iG} \end{pmatrix},\ \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_G \end{pmatrix},\ \mathbf{u}_i = \begin{pmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{iG} \end{pmatrix},\ \mathbf{Z}_i = \begin{pmatrix} \mathbf{z}_{i1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{z}_{i2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{z}_{iG} \end{pmatrix}$$
: GxL matrix of IV

<u>**Assumption SIV.1.**</u> $E(\mathbf{Z_i}'\mathbf{u}_i) = \mathbf{0}$

<u>**Assumption SIV.2.**</u> $rank E(\mathbf{Z_i}'\mathbf{X_i}) = K$ (rank condition for identification)

(If β is different across equations) under the above assumptions, 2SLS equation by equation is consistent (note that saying nothing on efficiency)

8.3 Generalized Method of Moments Estimation

8.3.1 A General Weighting Matrix

<u>GMM</u> : Since **L IVs are orthogonal to each u** ($E[\mathbf{Z_i}'\mathbf{u}_i] = E[\mathbf{Z_i}'(\mathbf{y}_i - \mathbf{X_i}\boldsymbol{\beta})] = \mathbf{0}$), obtain parameter estimates from the orthogonal (moment) condition (Obtain K parameter estimates from the system of **L equations**)

<u>**When L=K (just identified)**</u>

$$\hat{\boldsymbol{\beta}} = \left(N^{-1}\sum_{i=1}^{N}\mathbf{Z_i}'\mathbf{X_i}\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}\mathbf{Z_i}'\mathbf{y}_i\right) = (\mathbf{Z}'\mathbf{X})^{-1}(\mathbf{Z}'\mathbf{Y})$$
(System IV estimator)

From WLLN, this estimator is consistent

<u>**When L＞K (over identified)**</u>

We cannot obtain K parameters that satisfy all L (＞K) equations (cannot find unique betas satisfying $N^{-1}\sum_{i=1}^{N}\mathbf{Z_i}'(\mathbf{y}_i - \mathbf{X_i}\hat{\boldsymbol{\beta}}) = \mathbf{0}$) . Choose betas that makes $N^{-1}\sum_{i=1}^{N}\mathbf{Z_i}'(\mathbf{y}_i - \mathbf{X_i}\hat{\boldsymbol{\beta}})$ as close to zero as possible:

$$\min_{\boldsymbol{\beta}}\left[\sum_{i=1}^{N}\mathbf{Z_i}'(\mathbf{y}_i - \mathbf{X_i}\boldsymbol{\beta})\right]'\left[\sum_{i=1}^{N}\mathbf{Z_i}'(\mathbf{y}_i - \mathbf{X_i}\boldsymbol{\beta})\right]$$

<u>**GMM estimator**</u> (using weighting matrix)

$$\min_{\boldsymbol{\beta}}\left[\sum_{i=1}^{N}\mathbf{Z_i}'(\mathbf{y}_i - \mathbf{X_i}\boldsymbol{\beta})\right]'\hat{\mathbf{W}}\left[\sum_{i=1}^{N}\mathbf{Z_i}'(\mathbf{y}_i - \mathbf{X_i}\boldsymbol{\beta})\right],$$
$\hat{\mathbf{W}}$: LxL symmetric, positive semidefinite

$$=> \quad \hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{Z}\hat{\mathbf{W}}\mathbf{Z}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{Z}\hat{\mathbf{W}}\mathbf{Z}'\mathbf{Y}\right)$$
=> consistent? asymptotic normal?

<u>**Assumption SIV.3.**</u>**:** $\hat{\mathbf{W}} \xrightarrow{p} \mathbf{W}$ as $N \to \infty$, where W is a nonrandom, symmetric, LxL positive definite matrix.

Assume $\hat{\mathbf{W}}$ converges in probability to some positive definite matrix (Usually this is satisfied from WLLN because we choose $\hat{\mathbf{W}}$ as a function of sample mean)

<u>**Theorem 8.1**</u>**(consistency):** Under ASIV.1, 2, and 3, $\hat{\beta} \xrightarrow{p} \beta$ as $N \to \infty$.

<u>**Theorem 8.2**</u>**:** Under ASIV.1, 2, and 3, $\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} Normal[\mathbf{0}, A\operatorname{var}\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})]$ where $A\operatorname{var}\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = (\mathbf{C}'\mathbf{W}\mathbf{C})^{-1}\mathbf{C}'\mathbf{W}\Lambda\mathbf{W}\mathbf{C}(\mathbf{C}'\mathbf{W}\mathbf{C})^{-1}$, $\mathbf{C} \equiv E(\mathbf{Z}_i'\mathbf{X_i})$, $\Lambda \equiv E(\mathbf{Z_i}'\mathbf{u_i}\mathbf{u_i}'\mathbf{Z_i}) \equiv Var(\mathbf{Z_i}'\mathbf{u_i})$.

Note: estimates depend on **W** (**not in the limit because it is consistent estimator**)

8.3.2 The System 2SLS Estimator

Suppose $\hat{\mathbf{W}} = \left(N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{Z}_i\right)^{-1} = (\mathbf{Z'Z}/N)^{-1}$

$\hat{\boldsymbol{\beta}} = \left(\mathbf{X'Z(Z'Z)^{-1}Z'X}\right)^{-1}\left(\mathbf{X'Z(Z'Z)^{-1}Z'Y}\right)$ （**system 2SLS estimator**）

system 2SLS estimator is not necessarily efficient, but we can use it to calculate the initial value of optimal weighting matrix because it is consistent

8.3.3 The Optimal Weighting Matrix（How to choose $\hat{\mathbf{W}}$?）

＝＞**Choose $\hat{\mathbf{W}}$ so that the Asymptotic variance**

（$A\operatorname{var}\sqrt{N}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) = (\mathbf{C'WC})^{-1}\mathbf{C'W}\Lambda\mathbf{WC}(\mathbf{C'WC})^{-1}$） **becomes the smallest!**

<u>**Assumption SIV.4.**</u> $\mathbf{W} = \Lambda^{-1}$, where $\Lambda \equiv E(\mathbf{Z_i'u_iu_i'Z_i}) \equiv Var(\mathbf{Z_i'u_i})$.

1. ASIV4 ＝＞ $A\operatorname{var}\sqrt{N}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) = (\mathbf{C'}\Lambda^{-1}\mathbf{C})^{-1}$
2. $(\mathbf{C'WC})^{-1}\mathbf{C'W}\Lambda\mathbf{WC}(\mathbf{C'WC})^{-1} - (\mathbf{C'}\Lambda^{-1}\mathbf{C})^{-1}$ is positive semidefinite for any positive definite matrix W. **Therefore, asymptotic variance is the smallest when $\mathbf{W} = \Lambda^{-1}$!**

<u>**Theorem 8.3**</u> **:** Under ASIV.1, 2, 3, and 4, the resulting GMM estimator is efficient among all GMM estimators of the form $\hat{\boldsymbol{\beta}} = (\mathbf{X'Z\hat{W}Z'X})^{-1}(\mathbf{X'Z\hat{W}Z'Y})$.

<u>**How to estimate in practice?**</u>

<u>***Procedure 8.1***</u>

a. System 2SLS で、obtain an initial value of consistent estimator for $\boldsymbol{\beta}$ by system 2SLS （call it $\hat{\hat{\boldsymbol{\beta}}}$）

b. Obtain System 2SLS residual $\hat{\hat{\mathbf{u}}}_\mathbf{i} = \mathbf{y_i} - \mathbf{X_i}\hat{\hat{\boldsymbol{\beta}}}$

c. Obtain a consistent estimator of Λ : $\hat{\Lambda} = N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i'\hat{\hat{\mathbf{u}}}_i\hat{\hat{\mathbf{u}}}_i'\mathbf{Z}_i$

d. Estimate by GMM using $\hat{W} = \hat{\Lambda}^{-1} = \left(N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i'\hat{\hat{\mathbf{u}}}_i\hat{\hat{\mathbf{u}}}_i'\mathbf{Z}_i\right)^{-1}$

$$A\hat{\operatorname{var}}(\hat{\boldsymbol{\beta}}) = \left[(\mathbf{X'Z})\left(\sum_{i=1}^{N}\mathbf{Z}_i'\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i'\mathbf{Z}_i\right)^{-1}(\mathbf{Z'X})\right]^{-1},\quad \hat{\mathbf{u}}_\mathbf{i} = \mathbf{y_i} - \mathbf{X_i}\hat{\boldsymbol{\beta}}$$

8.3.4 The 3-stage Least Squares Estimator

3SLS estimator is a **GMM estimator** with $\hat{\mathbf{W}} = \left(N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i'\hat{\boldsymbol{\Omega}}\mathbf{Z}_i\right)^{-1} = [\mathbf{Z'}(\mathbf{I}_N\otimes\hat{\boldsymbol{\Omega}})\mathbf{Z}/N]^{-1}$, **(**$\hat{\boldsymbol{\Omega}} = N^{-1}\sum_{i=1}^{N}\hat{\hat{\mathbf{u}}}_i\hat{\hat{\mathbf{u}}}_i'$, $\hat{\hat{\mathbf{u}}}_\mathbf{i} = \mathbf{y_i} - \mathbf{X_i}\hat{\hat{\boldsymbol{\beta}}}$ **system 2SLS residuals). When is the 3SLS efficient?**

<u>**Assumption SIV.5.(“homoskedasticity”)**</u> $E(\mathbf{Z}_i'\mathbf{u}_i\mathbf{u}_i'\mathbf{Z}_i) = E(\mathbf{Z}_i'\boldsymbol{\Omega}\mathbf{Z}_i)$, where $\boldsymbol{\Omega} = E(\mathbf{u}_i\mathbf{u}_i')$.

<u>**Theorem 8.4**</u> **(Optimality of 3SLS):** Under ASIV.1, 2, 3, and 5, the 3SLS estimator is an optimal GMM estimator. Further, the appropriate estimator of asymptotic variance is;

$$A\hat{\operatorname{var}}(\hat{\boldsymbol{\beta}}) = \left[(\mathbf{X'Z})\left(\sum_{i=1}^{N}\mathbf{Z}_i'\hat{\boldsymbol{\Omega}}\mathbf{Z}_i\right)^{-1}(\mathbf{Z'X})\right]^{-1} = [\mathbf{X'Z}\{\mathbf{Z'}(\mathbf{I}_N\otimes\hat{\boldsymbol{\Omega}})\mathbf{Z}\}^{-1}\mathbf{Z'X}]^{-1}.$$

Under ASIV.5., 3SLS is efficient (GMM with optimal weighting matrix)

8.5 Testing Using GMM

Using Optimal GMM estimator $\hat{\boldsymbol{\beta}}$ and asymptotic variance $A\,\hat{\mathrm{var}}(\hat{\boldsymbol{\beta}}) = \left[(\mathbf{X}'\mathbf{Z})\left(\sum_{i=1}^{N}\mathbf{Z}_i{}'\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i{}'\mathbf{Z}_i\right)^{-1}(\mathbf{Z}'\mathbf{X})\right]^{-1}$,

t-test and Wald test are valid

<u>GMM distance test</u> (for linear constraints) $\left[\left(\sum_{i=1}^{N}\mathbf{Z}_i{}'\tilde{\mathbf{u}}_i\right)'\hat{\mathbf{W}}\left(\sum_{i=1}^{N}\mathbf{Z}_i{}'\tilde{\mathbf{u}}_i\right)-\left(\sum_{i=1}^{N}\mathbf{Z}_i{}'\hat{\mathbf{u}}_i\right)'\hat{\mathbf{W}}\left(\sum_{i=1}^{N}\mathbf{Z}_i{}'\hat{\mathbf{u}}_i\right)\right]/N \sim \chi^2_Q$

$\tilde{\mathbf{u}}_i$ **:**residual under restriction, $\hat{\mathbf{u}}_i$ **:**residual without restriction

<u>Test for nonlinear constraints</u> $H_0 : \mathbf{c}(\boldsymbol{\beta}) = \mathbf{0}$ **Qx1, C($\boldsymbol{\beta}$) : Jacobian matrix QxK**

Wald statistics: $\mathbf{c}(\hat{\boldsymbol{\beta}})'(\hat{\mathbf{C}}\hat{\mathbf{V}}\hat{\mathbf{C}}')^{-1}\mathbf{c}(\hat{\boldsymbol{\beta}}) \sim \chi^2_Q$

8.5.2 Testing Overidentification Restrictions

Test validity of additional IVs

<u>L>K (overidentified)</u>

$\left(N^{-1/2}\sum_{i=1}^{N}\mathbf{Z}_i{}'\hat{\mathbf{u}}_i\right)'\hat{\mathbf{W}}\left(N^{-1/2}\sum_{i=1}^{N}\mathbf{Z}_i{}'\hat{\mathbf{u}}_i\right) \sim \chi^2_{L-K}$ under $H_0 : E(\mathbf{Z}_i{}'\mathbf{u}_i) = \mathbf{0}$

（$\hat{\mathbf{W}}$: optimal weighting matrix）

If reject the null, need to reconsider the choice of IVs

8.6 More Efficient Estimation and Optimal Instruments (skip)

<u>How to choose IV(Z)?</u>

Basic solution: the more valid IV, the more efficient asymptotically

In particular, with heteroskedasticity it improves asymptotic efficiency (no efficiency improvement under homoskedasticity)

However, using too many IVs may cause the loss in finite property. In practice, we do not use infinite number of IVS.

Can we find a small set of optimal IVs rather than increasing the number of IVs to infinity?

<u>Assumption SIV.1’</u> $E(u_{ig} \mid \mathbf{z}_i) = 0$, $g = 1,\ldots,G$ for some vector z.

Z is exogenous in all equations

<u>Theorem 8.5</u> (Optimal Instruments): Under ASIV.1’ (and sufficient regularity conditions), the optimal choice of instruments is $Z_i^* = \boldsymbol{\Omega}(\mathbf{z}_i)^{-1}E(\mathbf{X}_i \mid \mathbf{z}_i)$, where $\boldsymbol{\Omega}(\mathbf{z}_i) = E(\mathbf{u}_i{}'\mathbf{u}_i \mid \mathbf{z}_i)$ provided that rank $E(\mathbf{Z}_i^*{}'\mathbf{X}_i) = K$.

PS 11

I. 8.1

II. 8.5